official

ALSTON&BIRD LLP

3201 Beechleaf Court, Suite 600
Raleigh, NC 27604-1062

919-862-2200
Fax: 919-862-2260

TELECOPY

PLEASE DELIVER AS SOON AS POSSIBLE

DATE: March 22, 2004

TO: U.S. Patent & Trademark Office/Art Unit 1638
Attn: Examiner David H. Kruse

FROM: Eric J. Kron

Re: Appl. No.: 09/954,950 Confirmation No.: 8514
Applicant(s): Mahajan et al
Filed: September 18, 2001
Art Unit: 1638
Title: RICE MLH1 ORTHOLOG AND USES THEREOF

Enclosed is an Amendment After Final (16 pages) w/Appendices 1-5.

Part II of II

NO. OF PAGES: 24
(Including cover page)

OPERATOR:

IF NOT RECEIVED PROPERLY, PLEASE NOTIFY ME IMMEDIATELY AT 919-862-2221

USER CODE:	KRONE	FAX NUMBER:	703-872-9306
CLIENT/MATTER:	035718/238971		
REQUESTED BY:	Nora Martinez	VOICE NUMBER:	571-272-0799

Mol Gen Genet (1999) 262: 239–249

ORIGINAL PAPER

J. Adé · F. Belzile · H. Philippe · M.-P. Doutriaux

Four mismatch repair paralogues coexist in *Arabidopsis thaliana*: *AtMSH2, AtMSH3, AtMSH6-1* and *AtMSH6-2*

Received: 23 February 1999 / Accepted: 5 May 1999

Abstract By using degenerate oligonucleotides based on the sequence homology between known MutS homologues, three *MSH* cDNAs belonging to the *MSH2*, *MSH3* and *MSH6* families, as defined in eukaryotes, have been isolated from *Arabidopsis thaliana* (ecotype Columbia). Genomic sequences for two of these genes (*AtMSH2* and *AtMSH6-2*) were also isolated and determined, whereas the genomic sequence of *AtMSH3* was obtained through the *Arabidopsis* sequencing project, as was the sequence of a second, distinct *AtMSH6* homologue (*AtMSH6-1*). Comparative analysis of the *AtMSH2* Landsberg *erecta* genomic sequence (reported here) and the previously described *AtMSH2* Columbia allele revealed several polymorphisms, including the presence of a small, transposon-like element in the 3′ untranscribed region of the former allele. *Arabidopsis* is the first organism to show such divergence of two *AtMSH6* genes; the divergence is strongly supported by sequence data and phylogenetic analysis. Southern analysis revealed that the three genes we have isolated exist as single copies, and genetic mapping indicated that *AtMSH2* and *AtMSH6-2* both reside on chromosome III. Finally, expression of these three genes could only be observed in suspensions of *A. thaliana* cells. Such a cell suspension divides actively after subculture, and the *AtMSH* genes are most strongly expressed at this stage.

Communicated by R. Devoret

J. Adé · F. Belzile
Département de Phytologie, 1243 Pavillon Marchand, Université Laval, Québec, Canada G1K 7P4

H. Philippe
Laboratoire de Biologie Cellulaire (CNRS URA 2227), Bât. 444, Université Paris-Sud, F-91405 Orsay Cedex, France

M.-P. Doutriaux (✉)
Institut de Biotechnologie des Plantes (CNRS UMR 8618), Bât. 630, Université Paris-Sud, F-91405 Orsay, France
e-mail: marie-pascale.doutriaux@ibp.u-psud.fr
Tel.: +33-169336349, Fax: +33-1-69336423

Key words *Arabidopsis thaliana* · Mismatch repair · Phylogenetic analysis

Introduction

The mismatch repair system (MMR) is essential for genetic stability, as it removes newly arising mutations from the genome and regulates recombination between related DNA sequences. MMR is responsible for the recognition and processing of mispaired bases that are spontaneously produced in the DNA as a consequence of replication errors, genetic recombination or deamination of 5-Me cytosines (Kolodner 1996). Repair of replication errors contributes to the conservation of the original information carried by the DNA. On the other hand, the regulation of the length of recombination intermediates (heteroduplexes) and whether or not they are edited by the MMR system define the degree to which recombination may occur between homologous but non-identical DNA sequences (Vulic et al. 1997). Recognition of mispaired bases in the heteroduplex region triggers the abortion of recombination events and prevents rearrangements between DNA sequences that are too divergent (Rayssiguier et al. 1989).

The MutHLS mismatch repair system in *Escherichia coli* is by far the best characterized. It derives its name from the three genes required to initiate MMR (MutH, MutL, MutS). The MutS protein is responsible for the detection of mismatches, and on binding it determines further processing and repair of mismatch-containing DNA molecules by the other components of the MMR. The MutH protein can recognize the newly replicated DNA strand, as it is transiently undermethylated at adenines in GATC sequences. Association of the MutL protein with MutS bound to the mismatch stimulates endonucleolytic cleavage of the unmethylated GATC sequence by MutH. Exonucleolytic degradation then proceeds to remove a stretch of up to 1000 bases around the mismatched base, followed by gap-repair synthesis and ligation of the correct DNA sequence (Modrich and

Lahue 1996). In the absence of a functional MMR system, bacteria show mutator or enhanced recombination proficiency phenotypes (Cox 1976; Feinstein and Low 1986; Rayssiguier et al. 1989).

Our current knowledge suggests that the general features of the bacterial MMR seem to be rather well conserved across all living organisms. Thus, mismatch repair activity or mismatch repair genes have been detected in a wide variety of eukaryotes (Modrich and Lahue 1996). Mismatch repair can be assayed following transfection of artificially constructed heteroduplex DNAs into yeast (Bishop et al. 1989; Kramer et al. 1989) and mammalian cells (Folger et al. 1985; Hare and Taylor 1985; Brown and Jiricny 1988), or after incubation in cell-free extracts from *Drosophila* (Holmes et al. 1990; Bhui-Kaur et al. 1998), *Xenopus* (Varlet et al. 1996) or human (Holmes et al. 1990; Thomas et al. 1991). This mismatch repair activity is abolished in all available mutants that are deficient for the mismatch repair functions (Parsons et al. 1993; Umar et al. 1994; Luhr et al. 1998). Finally, homologues of MutS and MutL have been isolated from eukaryotes, but their number suggests a higher level of complexity of the MMR system, or the involvement of more specialized processes. Whereas in bacteria the MutS proteins seem to belong to two different lineages (MutS-I and MutS-II as defined by Eisen 1998), which are not necessarily both present in every bacterial species, gene duplication and functional specialization have led to the divergence of many MutS homologues in eukaryotes. Six MutS homologues coexist in yeast: Msh2, Msh3 and Msh6 are involved in nuclear MMR, Msh4 and Msh5 participate in meiotic recombination and finally, Msh1 is involved in MMR in the mitochondria (Reenan and Kolodner 1992b; Ross-Macdonald and Roeder 1994; Hollingsworth et al. 1995; Marsischky et al. 1996). Msh2, Msh3 and Msh6 homologues have since been found in many different organisms, including mammals, *Drosophila*, *Neurospora* and *Arabidopsis* (for review, see Kolodner 1996). These eukaryotic *MSH* genes are classified as belonging to the MutS-I (*MSH1*, *MSH2*, *MSH3* and *MSH6*) or the MutS-II lineage (*MSH4* and *MSH5*) (Eisen 1998).

According to the current model for mismatch repair in eukaryotes Msh2 interacts with either Msh3 or Msh6 to form complexes with different recognition specifities: Msh2/3 complexes show a greater affinity for small insertions/deletions and Msh2/6 for single base mismatches (Kolodner 1996; Marsischky et al. 1996). This model is clearly supported by genetic and biochemical data (Reenan and Kolodner 1992b; Drummond et al. 1995; Palombo et al. 1995; Acharya et al. 1996; Marsischky et al. 1996; Genschel et al. 1998). Yeast *msh2* mutants exhibit a mutator phenotype reminiscent of the *mutS* phenotype in bacteria, as do the *msh3 msh6* double mutants (Marsischky et al. 1996). In humans, it is now well established that MMR deficiencies can lead to some hereditary cancer predisposition syndromes (Modrich and Lahue 1996). These cancers are associated with genetic instability, a phenotype that can be detected as an increased mutation rate in reporter genes or in tracts of short repeated DNA sequences. Such microsatellite instabilities presumably result from slippage of the replication machinery, which generates short insertions/deletions that would normally be recognized and repaired by the MMR. They are specifically observed in *msh2*, *msh3* or *msh6* tumor cells and in individuals with germline mutations in these genes (Modrich and Lahue 1996; Risinger et al. 1996; Akiyama et al. 1997; Miyaki et al. 1997). Genetic variability and cancer susceptibility are also dramatically increased in mice carrying null mutations in the *MSH2* or *MSH6* gene (de Wind et al. 1995; Reitmair et al. 1995; Edelmann et al. 1997).

As well as its role in surveillance of replication fidelity, the MMR is also involved in regulating genetic recombination between homologous but non-identical DNA sequences (Rayssiguier et al. 1989). If the outcome of recombination depends on the formation of a heteroduplex intermediate, the presence of mismatches in the heteroduplex makes it an obvious target for the MMR. A functional MMR system acting upon the mismatches can destabilize the heteroduplex, thus impeding recombination between homeologous DNA sequences. Studies in bacteria, yeast and mouse cells have all shown that mutations that affect components of the MMR can markedly increase the amount of recombination between divergent DNA sequences (Rayssiguier et al. 1989; Selva et al. 1995; de Wind et al. 1995; Datta et al. 1997).

Not much is known about mismatch repair in plants. Plant cell extracts from pea can repair mismatched oligonucleotides (Cerovic et al. 1991) and an *MSH2* homologue was recently isolated from *Arabidopsis thaliana* (Culligan and Hays 1997). With the aim of gaining insight into the role and activity of mismatch repair in plants, we have isolated homologues of *MSH2*, *MSH3* and *MSH6* from *A. thaliana*. Here, we provide a detailed characterization of these genes, which are expressed at detectable levels only in a mitotically active cell suspension derived from *Arabidopsis*.

Materials and methods

Growth of cell suspension

The cell suspension (ecotype Columbia) was initiated by Axelos et al. (1992) and is continuously propagated by weekly subculture (1.5 ml/25 ml) in Gamborg's B-5 basal medium (G-5893, Sigma), 30 g/l sucrose, 200 mg/l naphthalene acetic acid. The cell suspension is grown under agitation in a growth chamber with a 16 h photoperiod. Harvested plant material was stored at −70° C before extraction of RNA.

RNA isolation and Northern analysis

Total RNA was extracted from the cell suspension in the presence of Trizol (Gibco BRL) after homogenizing the cells in liquid N_2. Poly(A)$^+$ RNA was isolated using the Dynabeads mRNA Direct kit (Dynal). Poly(A)$^+$ RNA was fractionated in 1% agarose/formaldehyde gels after denaturation (Sambrook et al. 1989). Gels

were transferred onto Nylon Hybond N^+ membranes (Amersham) by capillary blotting. After hybridization to radiolabelled probes, the filters were washed in 0.1 × SCC, 0.1% SDS at 62° C and autoradiographed.

Genomic DNA isolation and Southern analysis

Genomic DNA was extracted from the cell suspension according to Dellaporta et al. (1983). Enzymatic digestion and electrophoresis of DNA was done using standard techniques. DNA was transferred onto Nylon Hybond N^+ membranes (Amersham) by capillary blotting. Genomic DNA sequences were isolated from a previously constructed *Sau*3AI partial genomic library (Doutriaux et al. 1998).

Radiolabelled probes

Radiolabelling of the probes with ^{32}P was carried out with the Stratagene Prime It II kit. Hybridization with ^{32}P-radiolabelled probes corresponding to the complete coding regions of the gene for translation elongation factor EF-1alpha from bean (pCHA0041; Axelos et al. 1989), *AtRAD51* (Doutriaux et al. 1998) and the 28S ribosomal RNA gene (Arabidopsis Biological Resource Center) were performed at 62° C according to Church and Gilbert (1984).

Reverse transcription and PCR

A 1-µg aliquot of total RNA was reverse transcribed using the MMLV reverse transcriptase by priming with random oligonucleotides in the presence of dNTPs. Using two different sets of degenerate oligonucleotides (each primer at 1 µM), PCR was performed using first-strand cDNA or genomic DNA in a final volume of 100 µl, in the presence of dNTPs (0.2 mM), 1 × PCR buffer and *Taq* polymerase (2 U). PCR parameters for oligonucleotides of set1 (touchdown PCR), were: three rounds of three cycles each (94° C for 1 min; 45° C, 41° C and 37° C for 2 min each, and 72° C for 1 min), followed by 35 cycles of 94° C for 30 s, 48° C for 30 s, and 72° C for 30 s, with a final step for 10 min at 72° C. For set2 oligonucleotides, PCR was carried out at 94° C for 5 min, followed by 30 cycles of 95° C for 40 s, 45° C for 1 min, and 72° C for 1 min. The amplification products At23 and At24 (obtained with set1), and S5 and S8 (set2) were subcloned and sequenced. Of these clones, At24 (654 bp, derived from genomic amplification) was homologous to *MSH2*, S5 (351 bp) was homologous to *MSH3*, and At23 (623 bp) and S8 (351 bp) were identical (except for the presence of introns in At23 – which was amplified from genomic DNA), and homologous to *MSH6*.

Isolation of *AtMSH2* cDNA

To obtain a cDNA clone for *AtMSH2*, ten pools of 10,000 clones each from library CD4-16 (ecotype Columbia, provided by the Arabidopsis Biological Resource Center) were plated on 15-cm petri dishes. The amplified phages were collected in 3 ml of SM buffer (10 mM NaCl, 1 mM $MgSO_4 \cdot 7H_2O$, 50 mM TRIS-HCl pH 7.5, 2% gelatin), of which 1 µl was used to perform PCR with the primers MSH2-1 and MSH2-2 that are specific for *AtMSH2*. One of the positive pools was used to generate ten pools of 1,000 clones each. PCR was used to identify positive pools of 1,000 phages from which two replicate filters were lifted. Two positive plaques were identified following hybridization with the At24 insert and in vivo excision of the insert (Stratagene) was used to obtain a plasmid version of one of the cloned cDNAs.

Isolation of the *AtMSH3* and *AtMSH6* cDNA sequences

Complete cDNA sequences were isolated using the procedure supplied with the Marathon cDNA amplification kit (Clontech). Briefly, double-stranded cDNA was produced by reverse transcription of 2 µg of poly(A)$^+$ RNA from the cell suspension culture of *Arabidopsis*. Adaptors were ligated on each side of the cDNA. The ligated cDNA was then used as a template for 5' and 3' RACE PCR reactions in the presence of primers specific for the adaptor on one side (AP1 and AP2), and specific for the targeted gene on the other side (see below), as defined from the previously isolated consensus regions S5 and S8. A 5' and a 3' fragment that overlap were thus produced for each gene.

Isolation of the complete coding sequence of *AtMSH3*

PCR performed on the ligated cDNA with primers 636 and AP1 for the 5' RACE PCR was followed by a second round of amplification with the nested primers AP2 and S525, which resulted in a 2720-bp DNA fragment. Another primer (S51) was designed that anneals closer to the 5' end and permitted the determination of 99 bp upstream of the ATG initiation codon. For the 3' RACE PCR, a first PCR reaction was performed with primers AP1 and 635, followed by a second round of amplification using the nested primers AP2 and S523, which produced a DNA fragment of 890 bp. Both DNA fragments were subcloned into pGEM-T and sequenced. Since PCR amplification using the Expand Long Template PCR system (Boehringer-Mannheim) produced errors in the sequence, new oligonucleotides were designed to re-isolate these sequences by PCR, using the high-fidelity DNA polymerase *Pfu*. PCR with primers 1S5 and S53 amplified a 1244-bp fragment (cloned into pUC18/SmaI). PCR with primers S52 and 2S5 amplified a 2104-bp fragment (cloned into pUC18/SmaI). These two clones were ligated after digestion with *Bam*HI, a unique site for which is present in the overlapping region. The complete reconstituted *AtMSH3* coding sequence is 3246 bp long.

Isolation of the complete coding sequence of *AtMSH6-2*

The same procedure allowed the isolation of the *AtMSH6-2* cDNA. For the 5' RACE PCR, primers 638 and AP1 allowed the amplification of a 2889-bp DNA fragment. Primer S81 helped define the 142 bp upstream of the ATG initiation codon. On the 3' side, RACE PCR was initially performed with primers S823 and AP1, and then with the nested primers 637 and AP2, to produce a 774-bp DNA fragment. As for *AtMSH3*, these fragments were cloned and sequenced. Due to PCR errors, re-isolation of this DNA sequence using the high-fidelity *Pfu* polymerase and the newly designed primers 1S8 and S83 (for the 5' side, 2182-bp clone 43 in pUC18/SmaI), and primers S82 and 2S8 (for the 3' side, 1379-bp clone 62 in pUC18/SmaI) was carried out. Clones 43 and 62 were digested with *Xmn*I, for which a unique site is present in the overlapping region, and ligated. The complete reconstituted *AtMSH6-2* coding sequence is 3330 bp long. An *AtMSH6-2* genomic sequence was also isolated from a genomic DNA library constructed from a partial *Sau*3AI digest of DNA from the *Arabidopsis* cell suspension. A stretch of 8062 bp was sequenced that included the entire *AtMSH6-2* gene, which was precisely colinear with the cDNA.

Oligonucleotides

Two sets of degenerate *MSH* primers were used. Set1 comprised MMR1 [5'-CGTGGATCCTCACIGGICCNAA(C/T)ATGGG-3'] and MMR2 (5'-GGTGAATTCGTGGAA(A/G)TGIGTNGC(A/G)AA-3'); Set2 (as in Reenan and Kolodner 1992a) consisted of MMR3 [5'-CTGGATCCACIGGICCIAA(C/T)ATG-3'] and MMR4 [5'-CTGGATCC(A/G)TA(A/G)TGIGTI(A/G)C(A/G)AA-3'].

As *AtMSH2* specific primers, MSH2-1 (5'-TCCACTTACATCCGCCAGGTTGATG-3'), MSH2-2 (5'-ATGCTCACATATAGCCCAAGCTAAACC-3') and MSH2-3 (5'-AAACTTGTGAGCTCGCTCTGCCCC-3') were used.

The *AtMSH3*-specific primers were 1S5 (5'-ATCCCGGGATGGGCAAGCAAAAGCAGCAGACGA-3'), 2S5 (5'-ATCCGGGTCAAAATGAACAAGTTGGTTTTAGTC-3'); S53 (5'-GACAAAGAGCGAAATGAGGCCCCTTGG-3'), S52 (5'-GCCACATCTGACTGTTCAAGCCCTCGC-3'), S51 (5'-GGATCGGGTACTGGGTTTTGAGTGTGAGG-3'); S525 (5'-AGGTTCTGATTATGTGTGACGCTTTACTTA-3'), S523 (5'-TCAGACAGTATCCAGCATGGCAGAAGTA-3'), 635 (5'-GCACGTGCTTGATGGTGTTTTCAC-3') and 636 (5'-TGCTAGTGCCTCTTGCAAGCTCAT-3').

The *AtMSH6-2* specific primers used were: 1S8 (5'-ATCCCGGGATGCAGCGCCAGAGATCGATTTTGT-3'), 2S8 (5'-ATCCCGGGTTATTTGGGAACACAGTAAGAGGATT-3'), S82 (5'-GCGTTCGATCATCAGCCTCTGTGTTGC-3'), S83 (5'-CGCTATCTATGGCTGCTTCGAATTGAG-3'), S81 (5'-CGTCGCCTTTAGCATCCCCTTCCTTCAC-3'); 637 (5'-GACAGCGTCAGTTCTTCAGAATGC-3'), 638 (5'-TCTCTACCAGGTGACGAAAAACCG-3') and S823 (5'-GCTTGGCGCATCTAATAGAATCATGACAGG-3').

Genetic mapping of *AtMSH2* and *AtMSH6-2*

Primers MSH2-1 and MSH2-3 were used to amplify a 1.3-kb segment of *AtMSH2* from the ecotypes Landsberg *erecta* and Columbia. A polymorphic *Mbo*I site was identified by sequence analysis and used to score 96 Recombinant Inbred (RI) lines resulting from a cross between Landsberg *erecta* and Columbia (Lister and Dean 1993). For *AtMSH6*, a RFLP between these two ecotypes was observed following digestion of genomic DNA with *Hin*dIII and hybridization with a PCR product of 2 kb. This polymorphism was scored in a subset (24) of the RI lines mentioned above. The MapMaker program (Lander et al. 1987) was used to determine the map positions of the *AtMSH2* and *AtMSH6-2* genes.

Phylogenetic analyses

Alignment of the sequences was carried out visually with the help of the ED program in the MUST package version 1.0 (Philippe 1993). Phylogenetic trees were constructed using maximum likelihood (ML), maximum parsimony (MP) and distance-based methods (Neighbor Joining, NJ) with the programs PROTML version 2.3 (Adachi and Hasegawa 1996), PAUP version 3.1 (Swofford 1993) and NJ in the MUST package version 1.0 (Philippe 1993), respectively. The distances were computed with the substitution model of Kimura (1983). MP trees were obtained by 100 random-addition heuristic search replicates, and ML trees by the Quick Add OTUs search, with the JTT model of amino acid substitution and retaining the 500 top-ranking trees (options -jf -q -n 500). Since it is important to take among-site rate variation into account in inferring phylogeny (Yang 1996), these 500 trees were further analysed with the PUZZLE program (Strimmer and von Haeseler 1996) as user trees with eight Gamma rate categories. Bootstrap proportions were calculated by the analysis of 1000 replicates for MP and NJ analysis. For ML analysis, bootstrap proportions were computed by using the RELL method (Kishino and Hasegawa 1989) owing to limitations on computing time.

Results

Isolation of the *AtMH2*, *AtMSH3* and *AtMSH6-2* cDNAs

Based upon a comparison of conserved amino acid sequences in known MutS-related proteins from various species, a set (set1) of degenerate oligonucleotides was designed; the second set (set2) used has been described previously (Reenan and Kolodner 1992a). PCR amplifications were performed using either *Arabidopsis* (ecotype Columbia) genomic DNA or first-strand cDNA as a template. This allowed the isolation of consensus regions for three potential homologues of *mutS*. At24 (654 bp), At23 (623 bp), S5 (351 bp) and S8 (351 bp) were cloned, and sequence analysis indicated that they were homologous, respectively, to *MSH2* (At24), *MSH3* (S5) and *MSH6* (At23, S8), three of the *MSH* genes previously described in yeast. After designing oligonucleotides specific for the genes of interest, two different approaches were taken in order to isolate their complete cDNA sequences. *AtMSH2* was isolated from a cDNA library, after successive rounds of selection of positive clones by PCR; *AtMSH3* and *AtMSH6-2* were isolated following the Marathon cDNA amplification procedure, which relies on 5' and 3' RACE-PCR.

The *AtMSH2* cDNA clone is 3039 bp long, and contains an ORF of 2811 nucleotides which is identical to that reported recently by Culligan and Hays (1997). The predicted protein is 937 amino acids long, with a predicted molecular weight of 105.5 kDa. The reconstituted *AtMSH3* sequence is 3553 bp long and contains a 3246-bp ORF with untranslated regions of 99 bp (5') and 144 bp (3') (EMBL/Genbank Accession No. AJ007791). The cDNA encodes a putative protein of 1081 amino acids, with a predicted molecular weight of 117.8 kDa. The *AtMSH6-2* sequence is 3701 bp long and contains an ORF encoding 1109 amino acids (predicted molecular weight 122.5 kDa); its coding region starts 141 bp from the 5' end and the polyA tail starts 106 bp downstream from the TAA stop codon (EMBL/Genbank Accession No. AJ007792). A short sequence (351 bp) that is identical to the *AtMSH6-2* consensus region has previously been described by Culligan and Hays (1997).

In the predicted protein sequences, the typical Msh functional domains can be found at the C terminal end (Fig. 1). Like other members of the MutS family, AtMsh2, AtMsh3 and AtMsh6-2 have the four motifs (A–D; see Fig. 1) characteristic of an NTP-binding domain, as defined by Gorbalenya and Koonin (1990) for the superfamily of UvrA-related proteins. The second conserved domain, containing the residues essential for the formation of the Helix-Turn-Helix structure (HTH; see Fig. 1) is also present in the *Arabidopsis* Msh proteins (Ohlendorf et al. 1983).

Genomic clones were isolated for both *AtMSH2* and *AtMSH6-2*. The *AtMSH2* genomic clone which we report here was isolated from the ecotype Landsberg *erecta* (GenBank Accession No. AF109243) and it shows several differences from the previously reported genomic clone of the Columbia allele (Culligan and Hays 1997; GenBank Accession No. AF003005). While the number and position of all 12 introns are conserved in both alleles, numerous polymorphisms are seen both in the coding and non-coding regions (see Fig. 2). Within the 13 exons, a total of 11 single-base substitutions was observed, of which six are neutral and five lead to a change in the amino acid sequence. None of these changes occurs at a position which is conserved among the eukaryotic

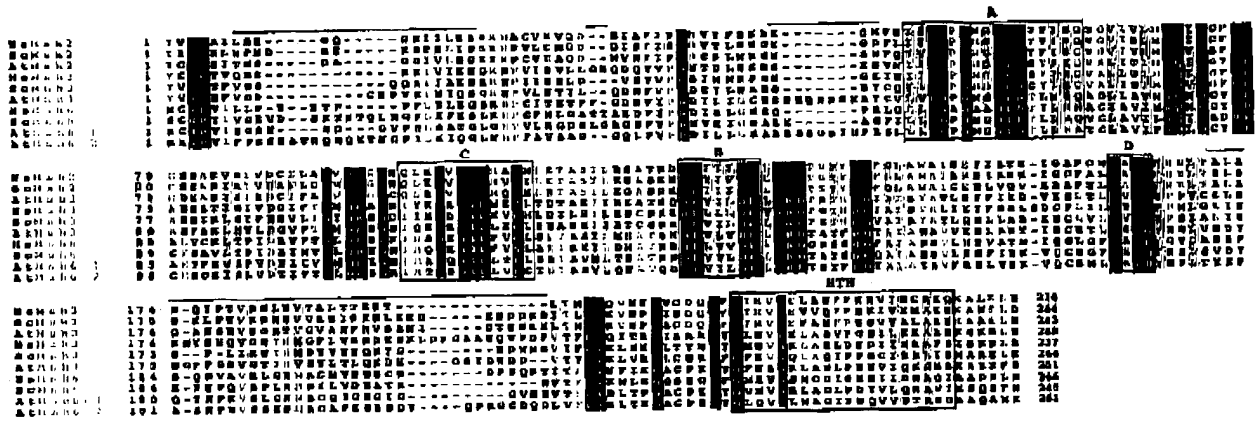
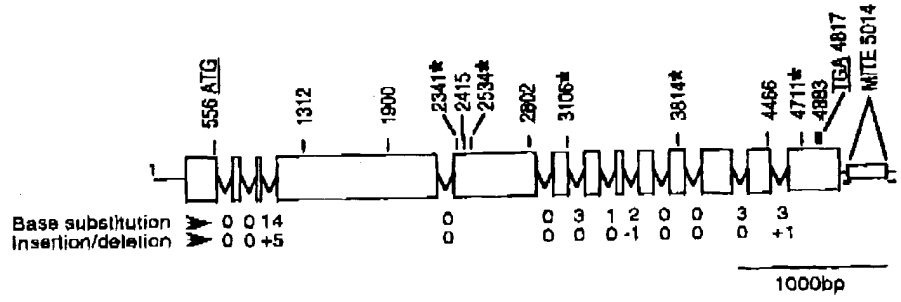


Fig. 1 Sequence alignment of the conserved C-terminal regions of Msh2, Msh3 and Msh6 from human, yeast and Arabidopsis. Gaps are indicated by *dashes*. Residues that are conserved or identical in all seven Msh sequences are enclosed in *light or dark gray boxes*, respectively. Letters in *blue*, *orange* and *green* indicate positions that are conserved in the aligned Msh2, Msh3 and Msh6 sequences, respectively. The four characteristic motifs (A, B, C and D) of the NTP-binding domain are delimited in *red* and the Helix-Turn-Helix (HTH) motif is delimited in *blue*. Highlighted sequences were omitted from consideration in the phylogenetic analysis and identity calculations

MSH2 genes. The most striking difference between the two alleles, however, is a 239-bp insertion located 196 bp after the stop codon in the 3′ untranscribed region of the Landsberg *erecta* allele. This insertion is flanked by a direct duplication of 5 bp and bears many but not all of the features of a miniature inverted-repeat transposable element (MITE), a class of small transposable elements recently reported in plants (Bureau and Wessler 1994b). This element differs from the *Emigrant* element, the only MITE reported to date in *Arabidopsis* (Casacuberta et al. 1998), and will be described in detail elsewhere (J. Ade and F. Belzile, unpublished).

A 8062-bp genomic region (Columbia ecotype) that encompasses the *AtMSH6-2* gene was also defined and revealed the presence of 16 introns within the sequenced region (EMBL/Genbank Accession No. AJ007792). The sequence of the genomic region of *AtMSH3* has been completed recently in the course of the *Arabidopsis* sequencing project. The 11 exons of this gene are found within a 5.5-kb stretch of the BAC clone M7J2 (GenBank Accession No. AL022197). Southern analysis of restriction digests of genomic *Arabidopsis* DNA with probes corresponding to the genomic consensus regions for the genes *AtMSH3* and *AtMSH6-2* indicates that they are single-copy genes and do not cross-hybridize (see Fig. 3). The sizes of the detected fragments always correlated exactly with their expected sizes, whenever these could be determined from available sequence information. Surprisingly, a fourth *MSH* gene was encountered in the course of the *Arabidopsis* genome sequencing project (ID ATAF1308, product name T10M13.8). Sequence comparisons indicate that this gene is related to

Fig. 2 Polymorphisms between the Landsberg *erecta* and Columbia alleles of *AtMSH2*. In this diagram, exons are shown as *open rectangles*, whereas introns are drawn as *v-shaped lines* between exons. The positions of the start (ATG) and stop (TGA) codons as well as that of each of the 11 polymorphisms (all single base substitutions) located within the coding region are indicated *above* the gene. Position 1 refers to the first base of the genomic sequence of this allele (GenBank accession AF109243). Substitutions that lead to a change in amino acid sequence are indicated by *asterisks*. The nature (number of base substitutions or length of insertion/deletion) of the polymorphisms located in introns is indicated *below* the diagram. A 239-bp miniature inverted-repeat transposable element (MITE-like) insertion (*hatched box*) flanked by a 5-bp duplication (–) at the insertion site was found in the 3′ region of the Landsberg *erecta* allele only



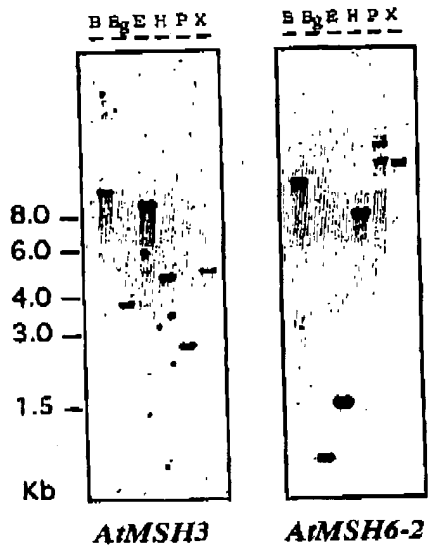


Fig. 3 Southern analysis of the genomic *AtMSH3* and *AtMSH6-2* loci. Total *Arabidopsis* DNA from the *Arabidopsis* cell suspension culture was digested with *Bam*HI (B), *Bgl*I (Bg), *Eco*RI (E), *Hind*III (H), *Pst*I (P) or *Xho*I (X). Positions of the size markers are shown on the left. The ^{32}P-radiolabelled probes used (S5 and S8, see Materials and methods) covered the consensus region of genes *MSH3* and *MSH6*, respectively

the *MSH6* family; since it was the first *AtMSH6* to be released in the databases we have named it *AtMSH6-1* in this study. Its complete genomic sequence includes 19 introns, of which only two coincide with introns in the *AtMSH6-2* sequence (data not shown).

Genetic mapping

The chromosomal positions of the genes *AtMSH2*, *AtMSH6-2* and *AtMSH3* were determined. In the case of *AtMSH2*, a CAPS marker was developed based on a polymorphic *Mbo*I site present in Landsberg *erecta* DNA but absent from Columbia DNA. This marker was used to follow the segregation of this locus among a population of 96 recombinant inbred lines (L. *erecta* × Columbia). *AtMSH2* was found to reside on the top arm of *Arabidopsis* chromosome III, 1.1 cM from locus m105. For *AtMSH6-2*, a RFLP mapping approach was used on a subset of 24 RI lines and this indicated that *AtMSH6-2* was also located on chromosome III and cosegregates with *ABI3*. For *AtMSH3*, both CAPS and RFLP approaches were unsuccessful, owing to a lack of detectable polymorphism between the mapping ecotypes. However, the location of the recently sequenced BAC clone (M7J2) containing the *AtMSH3* gene indicates that it maps on the top of chromosome IV (closest marker PG19). *AtMSH6-1* also resides on chromosome IV, based on the mapping of BAC clone T10M13 (closest marker GT148).

Amino acid sequence comparisons

Initially, the sequence alignment was restricted to the conserved region which comprises the four NTP-binding domains in the C terminal region of the proteins (roughly 250 amino acids, see Fig. 1). As a general observation, the deduced sequences of the different AtMsh proteins are more similar to their human counterparts than to the yeast homologues. In the conserved region, AtMsh2 is 71% identical to the human Msh2 protein, AtMsh3 is 59% identical to the human Msh3 and AtMsh6-1 and AtMsh6-2 are, respectively, 55% and 54% identical to the human Msh6 (see Table 1). While these levels of identity are the highest observed, the *Arabidopsis* consensus sequences also resemble their respective orthologues (i.e. members of the same MutS family in other species) more than any paralogous Msh family members (i.e. members of other MutS families in the same species). The two *Arabidopsis* Msh6 amino acid sequences differ from each other, but still resemble each

Table 1 Percentage identity between Msh2, Msh3 and Msh6 sequences from humans, yeast and Arabidopsis

Protein	Species[a]	Percent identity[b]									
Msh2	Hs	–									
	Sc	73 (41)	–								
	At	71 (40)	69 (36)	–							
Msh3	Hs	47 (23)	47 (21)	45 (21)	–						
	Sc	43 (23)	42 (23)	40 (23)	52 (31)	–					
	At	44 (22)	46 (22)	50 (23)	59 (36)	50 (28)	–				
Msh6	Hs	46 (22)	45 (21)	47 (23)	46 (24)	44 (23)	46 (22)	–			
	Sc	44 (20)	42 (20)	41 (21)	41 (24)	38 (24)	42 (24)	51 (20)	–		
	At-1	45 (21)	43 (21)	43 (21)	50 (25)	40 (23)	44 (23)	55 (31)	48 (29)	–	
	At-2	46 (20)	44 (22)	46 (21)	45 (24)	38 (17)	43 (19)	54 (29)	49 (28)	58 (29)	–
		Hs	Sc	At	Hs	Sc	At	Hs	Sc	At-1	At-2
			Msh2			Msh3			Msh6		

[a] Hs, human; Sc, yeast; At, Arabidopsis
[b] Values were calculated based on the 197 alignable residues as described in Fig. 1. Values in *parentheses* were calculated for the full-length proteins aligned using CLUSTALW

other (58% identity) more than the human or yeast Msh6. The three proteins described in this study are clearly not members of the Msh1 or Msh4 and Msh5 families, the mitochondrial and meiotic MutS homologues, respectively. Comparison of the full-length proteins corroborates these observations, although the alignment with CLUSTALW over their complete sequences should be considered less accurate since no further manual refinement was done. This analysis shows that AtMsh2 is 40% identical (61% similar) to the entire human Msh2 protein; AtMsh3 is 36% identical (54% similar) to the human Msh3 and AtMsh6-2 is 29% identical (44% similar) to its human counterpart (human Msh6 and AtMsh6-1 are 31% identical and 48% similar to each other). In all instances, the levels of identity and/or similarity mentioned above are higher than for any other combination of compared proteins presented in Table 1. Furthermore, along the aligned protein sequences, some amino acid motifs or positions are found to be specifically conserved within each of the three Msh families and the *Arabidopsis* proteins also present these specific patterns (data not shown).

A phylogenetic study was performed using 197 unambiguously aligned amino acids within or around the consensus region at the C-terminal end (as in Fig. 1). All available Msh2, 3 and 6 sequences were analysed using a maximum likelihood (ML) method which takes into account the differential rate of amino acid substitutions among different sites in a protein (through the use of a gamma law). The pattern of the tree confirms the classification established based on degree of identity (Fig. 4). The three groups consisting of Msh2, Msh3 or Msh6 homologues are distinctly defined, and the *Arabidopsis* Msh sequences we isolated are firmly assigned to their respective groups. The evolutionary rate is lowest for Msh2, and highest for Msh3 and Msh6 proteins. In all three Msh subgroups, homologues from plants and animals (except for the *Drosophila* Msh2) tend to group together and are separated from Msh proteins from fungi. The occurrence of two intraspecies Msh6 homologues that evolve independently may be restricted to *Arabidopsis* (or plants), since divergence of these two genes seems to have occurred after the divergence of plants and animals.

Conservation of a few intron positions between the two *AtMSH6* genes reinforces this observation: if only strictly aligned intron positions are taken into account, two introns are found at exactly the same positions in both *AtMSH6* genes (introns 5 and 14 of *AtMSH6-2*/ introns 7 and 15 of *AtMSH6-1*); one of these sites also harbours an intron in the *AtMSH3* gene (intron 14 of *AtMSH6-2*/intron 15 of *AtMSH6-1*/intron 10 of *AtMSH3*). None of these intron positions is shared by *AtMSH2* or any other *MSH6* (data not shown).

Expression studies

Expression of the different *AtMSH* genes was assessed by Northern analysis performed with poly(A)$^+$ RNA (see Fig. 5). The size and the low expression level of these genes made it necessary to use poly(A)$^+$ RNA, as their transcripts migrate in the same region as the 28S RNA, which makes the signal too diffuse to be detectable with confidence by autoradiography. Since the *AtMSH* genes are very poorly expressed in plant tissues (data not shown), we took advantage of an *A. thaliana* cell suspension (Axelos et al. 1992). This cell suspension is mitotically active: the cells grow exponentially for the first 5 days following inoculation, before entering the stationary phase; then the number of growing cells, as measured by their ability to form protoplasts, starts to decrease (data not shown). Northern analysis identified mRNAs of approximately 3.4 kb for *AtMSH2*, 3.5 kb for *AtMSH3* and 3.7 kb for *AtMSH6-2*, in accordance with the sizes predicted from the isolated cDNAs (Fig. 5). On day 2, when the cells are in the early exponential growth phase, the *AtMSH6-2* transcript is expressed at a higher level than at day 8; the same is true for *AtMSH2* and *AtMSH3*, albeit to a lesser extent.

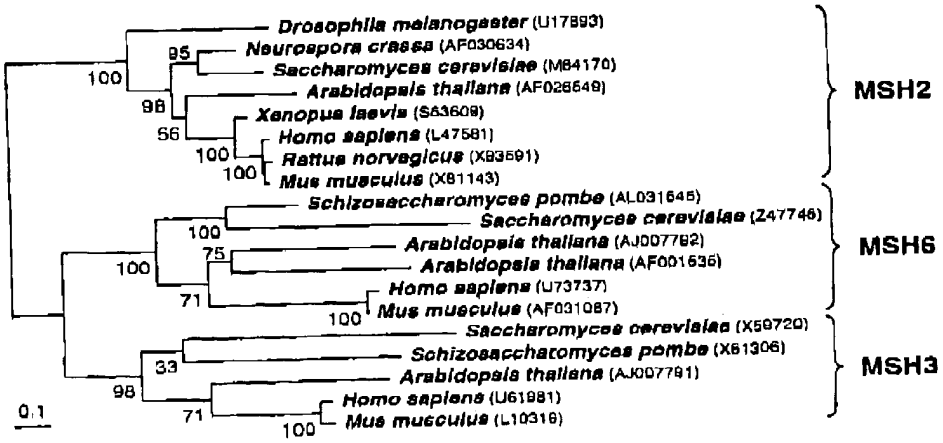


Fig. 4 Phylogenetic analysis of the 197 aligned amino acids from the conserved region of all available MSH2, MSH3 and MSH6 sequences. In such a tree the length of the horizontal branches is such that the evolutionary distance between two proteins is proportional to the total length of the horizontal branches that connect them (vertical branch lengths are arbitrary). Bootstrap values are shown at the nodes and the side bar represents 10% sequence divergence

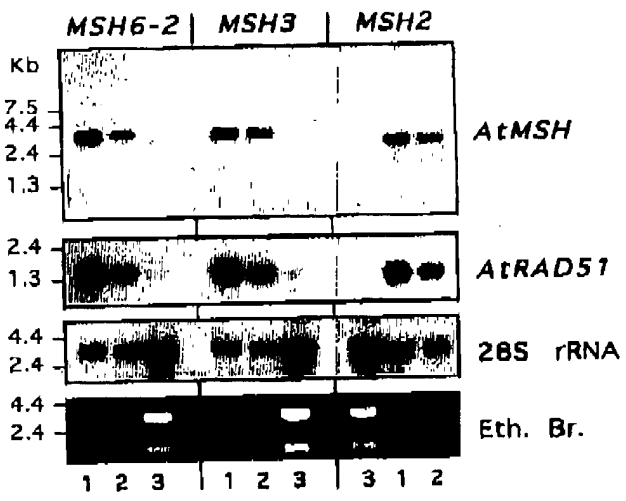


Fig. 5 Northern analysis of the *MSH* gene expression in *Arabidopsis* suspension cultures. Aliquots (2 μg) of poly(A)$^+$ RNA were loaded in lanes 1 and 2; 1 μg of total RNA was loaded in lane 3. RNA was extracted from the *Arabidopsis* cell suspension cultures 2 days (lane 1) and 8 days (lanes 2 and 3) after subculture. Northern hybridization was carried out with the ^{32}P-labelled 5′ regions of the different *MSH* genes (excluding the consensus region), and with complete *AtRAD51* and bean translational elongation factor (EF) cDNAs and 28S rRNA sequences. The bottom panel shows the ethidium bromide staining pattern

These data can be correlated with *AtRAD51* expression, which is higher on day 2 than day 8, as expected (Doutriaux et al. 1998). The probes used covered the 5′ regions of the three genes, outside the consensus regions, and were chosen so that they would not cross-hybridize with the different *MSH* genes. This is confirmed by the fact that a single band of the expected size was detected with each probe.

Discussion

On the basis of their relationship with *MSH* genes from other species, we have isolated three *Arabidopsis* homologues of the *mutS* gene, which is known to be essential for the repair of DNA mismatches in *E. coli*. Taking into account another *MSH* gene encountered during the course of the *Arabidopsis* genome sequencing project, four *MSH* genes are now known in this plant species: *AtMSH2*, *AtMSH3* and two *AtMSH6* (*-1* and *-2*). These four genes share sequence characteristics with all *MSH* family members. The lengths and molecular weights of the predicted proteins are similar to those of other MutS homologues. In the C-terminal region, two highly conserved motifs – the NTP-binding domain and a Helix-Turn-Helix domain – are found. From sequence comparisons, we classify them as belonging to the Msh2, Msh3 and Msh6 families. At the genomic level, the three *MSH* genes we isolated are unique and are detected as single bands following digestion of genomic DNA with restriction enzymes that do not recognize any sites in the probe. Nevertheless, another *MSH6* homologue exists in *Arabidopsis*: *MSH6-1* was detected during the systematic sequencing of *Arabidopsis* chromosome IV (Johnson et al. 1997). This gene differs from *AtMSH6-2* in terms of sequence, chromosomal location, and intron distribution; furthermore, it is not detected with a probe that includes the conserved region of *AtMSH6-2*.

Sequence comparisons between the conserved C-terminal regions of the Msh proteins, as well as their complete sequences, clearly allow us to designate the four *AtMSH* genes of *Arabidopsis* as belonging to the *MSH2*, *MSH6* or *MSH3* family. Such comparisons also led to other important observations. It appears that the levels of identity are much higher among the Msh2 orthologues than is the case among the Msh3 or Msh6 orthologues. It is commonly believed that the more interactions a protein is involved in, the lower its rate of evolution (Dickerson 1971). As a functional mismatch repair system relies on Msh2 binding either to Msh3 or Msh6, followed by an interaction with the Mlh1/Pms1 complex, Msh2 interacts with at least three proteins while Msh3 and Msh6 only bind to Msh2 (Prolla et al. 1994; Acharya et al. 1996). Two-hybrid experiments have also identified PCNA, Exo1 and components of the nucleotide excision repair pathway as partners of the yeast or human Msh2 (Umar et al. 1996; Tishkoff et al. 1997; Bertrand et al. 1998; Gu et al. 1998). Such an experimental approach has not been reported for Msh3 and Msh6 and therefore we cannot exclude the possibility that other proteins interact with these gene products. However, all indications suggest that Msh2 lies at the center of a complex protein network; this might therefore more severely restrict the sequence fluctuations permissible for this protein.

A phylogenetic study including all available eukaryotic Msh (2, 3 and 6) sequences confirmed the previous assignment of each of the four putative Msh proteins from *Arabidopsis* to the three different Msh families. The analysis of the phylogeny of the Msh proteins is complicated by the heterogeneity of evolutionary rates, which can lead to artefacts in tree construction – all the more so when the number of nucleotides used is low. The difficulty in reconstructing Msh phylogeny is illustrated by the fact that the monophyly of fungal Msh6 sequences is only recovered when among-site rate variation is taken into account (data not shown). Variation in evolutionary rates is not only observed between paralogues but also between species within each group of paralogues (see for example the branch lengths for fungal sequences in Fig. 4). The most likely artefact is the long branch attraction phenomenon (Felsenstein 1978), which generally results in the incorrect early emergence of fast-evolving sequences (Philippe and Laurent 1998). For instance, the Msh2 sequence of *Drosophila*, which emerges at the base of this group – far from other Metazoa – is very likely to be misplaced

because of this phenomenon. Similarly, although fungi are known to be the closest relatives of animals (Baldauf and Palmer 1993), phylogenies based on Msh have always found that animals are closely related to plants and not to fungi, which could be due to a higher rate of evolution of MMR proteins in fungi. Interestingly the same observation, i.e. an increased evolutionary rate in fungi and *Drosophila*, has previously been reported in a phylogenetic analysis of Rad51 (Yeager Stassen et al. 1997). Finally, the finding that two intron positions are common to both *Arabidopsis MSH6* genes and that one of these is also coincident with an intron position in *AtMSH3* may argue in favour of the relatedness of these two families.

The occurrence within the same species of two Msh6 homologues that evolve independently is unexpected, and may be restricted to *Arabidopsis* (or plants) since divergence of these two genes occurred after the separation of plants and animals. Nevertheless, one may speculate about the situation in other eukaryotes. As its genome has been completely sequenced, there is no doubt about the uniqueness of *MSH6* in *S. cerevisiae*. Although a single *MSH6* gene has been described for human and mouse, this cannot be considered definitive as long as the human genome has not been totally sequenced. Despite the considerable effort that has been devoted to the identification of all human *MSH* genes, *MSH4* and *MSH5*, the meiotic *mutS* homologues, were discovered only recently (Paquis-Flucklinger et al. 1997; Her and Doggett 1998). Except for the particular case of *Sarcophyton glaucum* (Pont-Kingdon et al. 1998), no *MSH1* gene has yet been found in any higher eukaryote, but, in the view of the rate and spectrum of mitochondrial mutations in human cells, its existence remains questionable (Khrapo et al. 1997). In fact, the single-copy nature of the *MSH6* genes is strongly supported by the persistence of a specific and similar phenotype when Msh6 is defective in either *S. cerevisiae* or mammals (Marsichsky et al. 1996; Edelmann et al. 1997). Phylogenetic analysis also favour the idea that only plants have two Msh6 proteins that evolve independently. In the absence of expression or functional studies, it is clear that we cannot yet conclude that an AtMsh6-1 protein is in fact active in *Arabidopsis*. However, the phylogenetic analysis suggests that AtMsh6-1 is a functional protein, otherwise it would be expected to have picked up mutations in the known functional domains. Whether Msh6-2 and AtMsh6-1 coexist in the same tissues, are functionally redundant or have acquired different specialized functions will have to be assessed in the future.

The three *MSH* genes we describe are expressed in *Arabidopsis*: cDNA clones were successfully obtained and mRNAs specific to each gene were detected in Northern blot experiments. The *AtMSH2*, *AtMSH3* and *AtMSH6-2* transcripts differ in size and their estimated sizes correlate with the lengths of the cDNA sequences. All three genes are expressed in a cell suspension derived from *A. thaliana* – at slightly higher levels in the exponential growth phase than in the stationary phase. We also find a much higher level of *AtRAD51* transcripts in the cells on day 2 than on day 8; *AtRAD51* has been shown previously to be regulated in S-phase (Doutriaux et al. 1998). Precise assessment of the phase of induction in the cell cycle would require cell synchronization. In *S. cerevisiae*, induction in early S-phase has been described for the *MSH2* and *MSH6* genes, while *MSH3* transcription was found to be constitutive during the cell cycle (Kramer et al. 1996). In *E. coli*, MutS was also found to be depleted in stationary-phase cultures (Feng et al. 1996). Overall, these data support the idea that *MSH* genes are expressed at a time when cells are dividing actively and thus replicating their DNA.

Acknowledgements We thank Catherine Bergounioux, Najat Takvorian and Michael Hodges for valuable comments on the manuscript and Roland Boyer for photographic work. This work was supported in part by CNRS and Université Paris XI (M.-P. D. and H. P.), Biogemma and Danone (M.-P. D.), and in part by NSERC (J. A. and F. B.) (Canada). J. A. also wishes to acknowledge a fellowship from the Programme Canadien de Bourses de la Francophonie.

References

Acharya S, Wilson T, Gradia S, Kane MF, Guerrette S, Marsischky GT, Kolodner R, Fishel R (1996) hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. Proc Natl Acad Sci USA 93:13629–13634

Adachi J, Hasegawa M (1996) MOLPHY: programs for molecular phylogenetics. Institute of Statistical Mathematics, Tokyo

Akiyama Y, Sato H, Yamada T, Nagasaki H, Tsuchiya A, Abe R, Yuasa Y (1997) Germ-line mutation of the *hMSH6/GTBP* gene in an atypical hereditary nonpolyposis colorectal cancer kindred. Cancer Res 57:3920–3923

Axelos M, Bardet C, Liboz T, Le Van Thai A, Curie C, Lescure B (1989) The gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1 alpha: molecular cloning, characterization and expression. Mol Gen Genet 219:106–112

Axelos M, Curie C, Mazzolini L, Bardet C, Lescure B (1992) A protocol for transient gene expression in *Arabidopsis thaliana* protoplasts isolated from cell suspension culture. Plant Physiol Biochem 30:1–6

Baldauf SL, Palmer JD (1993) Animals and fungi are each other's closest relatives. Science 257:74–76

Bertrand P, Tishkoff DX, Filosi N, Dasgupta R, Kolodner RD (1998) Physical interaction between components of DNA mismatch repair and nucleotide excision repair. Proc Natl Acad Sci USA 95:14278–14283

Bhui-Kaur A, Goodman MF, Tower J (1998) DNA mismatch repair catalyzed by extracts of mitotic, postmitotic, and senescent *Drosophila* tissues and involvement of *mei-9* gene function for full activity. Mol Cell Biol 18:1436–1443

Bishop DK, Andersen J, Kolodner RD (1989) Specificity of mismatch repair following transformation of *Saccharomyces cerevisiae* with heteroduplex plasmid DNA. Proc Natl Acad Sci USA 86:3713–3717

Brown TC, Jiricny J (1988) Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. Cell 54:705–711

Bureau TE, Wessler SR (1994) *Stowaway*: a new family of inverted-repeat elements associated with the genes of both monocotyledonous and dicotylenodous plants. Plant Cell 6:907–916

Casacuberta E, Casacuberta JM, Puigdomenech P, Monfort A (1998) Presence of a miniature inverted-repeat transposable element (MITEs) in the genome of *Arabidopsis thaliana*: characterization of the *Emigrant* family of elements. Plant J 16:79–85

Cerovic G, Bozin D, Dimitrijevic B (1991) Mismatch-specific DNA breakdown in nuclear extract from tobacco (*Nicotiana tabacum*) callus. Plant Mol Biol 17:887–894

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995

Culligan KM, Hays JB (1997) DNA mismatch repair in plants. An *Arabidopsis thaliana* gene that predicts a protein belonging to the MSH2 subfamily of eukaryotic MutS homologs. Plant Physiol 115:833–839

Cox EC (1976) Bacterial mutator genes and the control of spontaneous mutation. Annu Rev Genet 10:135–156

Datta A, Hendrix M, Lipsitch M, Jinks-Robertson S (1997) Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. Proc Natl Acad Sci USA 94:9757–9762

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Report 1:19–21

Dickerson RE (1971) The structures of cytochrome *c* and the rates of molecular evolution. J Mol Evol 1:26–45

Doutriaux MP, Couteau F, White C (1998) Isolation and characterisation of the RAD51 and DMC1 homologs from *Arabidopsis thaliana*. Mol Gen Genet 257:283–291

Drummond JT, Li GM, Longley MJ, Modrich P (1995) Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. Science 268:1909–1912

Edelmann W, Yang K, Umar A, Heyer J, Lau K, Fan K, Liedtke W, Cohen PE, Kane MF, Lipford JR, Yu N, Crouse GF, Pollard JW, Kunkel T, Lipkin M, Kolodner R, Kucherlapati R (1997) Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. Cell 91:467–477

Eisen JA (1998) A phylogenomic study of the MutS family of proteins. Nucleic Acids Res 26:4291–4300

Feinstein SI, Low KB (1986) Hyper-recombining recipient strains in bacterial conjugation. Genetics 113:13–33

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–410

Feng G, Tsui HCT, Winkler ME (1996) Depletion of the cellular amounts of the MutS and MutH methyl-directed mismatch repair proteins in stationary-phase *Escherichia coli* K-12 cells. J Bacteriol 178:2388–2396

Folger KR, Thomas K, Capecchi MR (1985) Efficient correction of mismatched bases in plasmid heteroduplexes injected into cultured mammalian cell nuclei. Mol Cell Biol 5:70–74

Genschel J, Littman SJ, Drummond JT, Modrich P (1998) Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. J Biol Chem 273:19895–19901

Gorbalenya AE, Koonin EV (1990) Superfamily of UvrA-related NTP-binding proteins: implications for rational classification of recombinational/repair systems. J Mol Biol 213:583–591

Gu L, Hong Y, McCulloch S, Watanabe H, Li GM (1998) ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. Nucleic Acids Res 26:1173–1178

Hare JT, Taylor JH (1985) One role for DNA methylation in vertebrate cells is strand discrimination in mismatch repair. Proc Natl Acad Sci USA 82:7350–7354

Her C, Doggett NA (1998) Cloning, structural characterization, and chromosomal localization of the human orthologue of the *Saccharomyces cerevisiae MSH5* gene. Genomics 52:50–61

Hollingsworth NM, Ponte L, Halsey C (1995) MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in *Saccharomyces cerevisiae* but not mismatch repair. Genes Dev 9:1728–1739

Holmes J Jr, Clark S, Modrich P (1990) Strand-specific mismatch correction in nuclear extracts of human and *Drosophila melanogaster* cell lines. Proc Natl Acad Sci USA 87:5837–5841

Johnson AF, de la Bastide M, Lodhi M, Hoffman J, Hasegawa A, Gnoj L, Gottesman T, Granat S, Hameed A, Kaplan N, Schutz K, Shohdy N, Van Keuren K, Parnell L, Dedhia N, Martienssen R, McCombie W (1997) The sequence of the *Arabidopsis thaliana* BAC T10M13. GenBank Accession No. AF001308

Khrapo K, Coller HA, André PC, Li XC, Hanekamp JS, Thilly WG (1998) Mitochondrial mutational spectra in human cells and tissues. Proc Natl Acad Sci USA 94:13798–13803

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J Mol Evol 29:170–179

Kolodner RD (1996) Biochemistry and genetics of eukaryotic mismatch repair genes. Genes Dev 10:1433–1442

Kramer W, Fartmann B, Ringbeck EC (1996) Transcription of *mutS*-and *mutL*-homologous genes in *Saccharomyces cerevisiae* during the cell cycle. Mol Gen Genet 252:275–283

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Noewburn L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J 4:745–750

Luhr B, Scheller J, Meyer P, Kramer W (1998) Analysis of in vivo correction of defined mismatches in the DNA mismatch repair mutants *msh2*, *msh3* and *msh6* of *Saccharomyces cerevisiae*. Mol Gen Genet 257:362–367

Marsischky GT, Filosi N, Kane MF, Kolodner R (1996) Redundancy of *Saccharomyces cerevisiae* MSH3 and MSH6 in MSH2-dependent mismatch repair. Genes Dev 10:407–420

Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, Igari T, Koike M, Chiba M, Mori T (1997) Germline mutation of *MSH6* as the cause of hereditary nonpolyposis colorectal cancer. Nature Genet 17:271–272

Modrich P, Lahue R (1996) Mismatch repair in replication fidelity, genetic recombination and cancer biology. Annu Rev Biochem 65:101–133

Ohlendorf DH, Anderson WF, Matthews BW (1983) Many gene-regulatory proteins appear to have a similar alpha-helical fold that binds DNA and evolved from a common precursor. J Mol Evol 19:109–114

Palombo F, Gallinari P, Iaccarino I, Lettieri T, Hughes M, D'Arrigo A, Truong O, Hsuan JJ, Jiricny J (1995) GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. Science 268:1912–1914

Paquis-Flucklinger V, Santucci-Darmanin S, Paul R, Turc-Carel C, Desnuelle C (1997) Cloning and expression analysis of a meiosis-specific MutS homolog: the human *MSH4* gene. Genomics 44:188–194

Parsons R, Li GM, Longley MJ, Fang WH, Papadopoulos N, Jen J, de la Chapelle A, Kinzler KW, Vogelstein B, Modrich P (1993) Hypermutability and mismatch repair deficiency in RER+ tumor cells. Cell 75:1227–36

Philippe H (1993) MUST, a computer package of Management Utilities for Sequences and Trees. Nucleic Acids Res 21:5264–5272

Philippe H, Laurent J (1998) How good are deep phylogenetic trees? Curr Opin Genet Dev 8:616–623

Pont-Kingdon G, Okuda NA, Macfarlane JL, Beagley CT, Watkins-Sims CD,Cavalier-Smith T, Clark-Walker GD, Wolstenholme DR (1998) Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. J Mol Evol 46:419–431

Prolla TA, Pang Q, Alani E, Kolodner RD, Liskay RM (1994) MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. Science 265:1091–1093

Rayssiguier C, Thaler DS, Radman M (1989) The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature 342:396–401

Reenan RA, Kolodner RD (1992a) Isolation and characterization of two *Saccharomyces cerevisiae* genes encoding homologs of the bacterial HexA and MutS mismatch repair proteins. Genetics 132:963–973

Reenan RA, Kolodner RD (1992b) Characterization of insertion mutations in the *Saccharomyces cerevisiae MSH1* and *MSH2* genes: evidence for separate mitochondrial and nuclear functions. Genetics 132:975–985

Reitmair AH, Schmits R, Ewel A, Bapat B, Redston M, Mitri A, Waterhouse P, Mittrücker HW, Wakeham A, Liu B, Thomason A, Griesser H, Gallinger S, Ballhausen WG, Fishel R, Mak TW (1995) MSH2-deficient mice are viable and susceptible to lymphoid tumours. Nature Genet 11:64–70

Risinger JI, Umar A, Boyd J, Berchuck A, Kunkel TA, Barrett JC (1996) Mutation of *MSH3* in endometrial cancer and evidence for its functional role in heteroduplex repair. Nature Genet 14:102–105

Ross-Macdonald P, Roeder GS (1994) Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. Cell 79:1069–1080

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Selva EM, New L, Crouse GF, Lahue RS (1995) Mismatch correction acts as a barrier to homeologous recombination in *Saccharomyces cerevisiae*. Genetics 139:1175–1188

Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Swofford DL (1993) PAUP: Phylogenetic Analysis Using Parsimony. Distributed by the Illinois Natural History Survey, Champaign

Thomas DC, Roberts JD, Kunkel TA (1991) Heteroduplex repair in extracts of human HeLa cells. J Biol Chem 266:3744–3751

Tishkoff DX, Boerger AL, Bertrand P, Filosi N, Gaida GM, Kane MF, Kolodner RD (1997) Identification and characterization of *Saccharomyces cerevisiae EXO1*, a gene encoding an exonuclease that interacts with MSH2. Proc Natl Acad Sci USA 94:7487–7492

Umar A, Boyer JC, Thomas DC, Nguyen DC, Risinger JI, Boyd J, Ionov Y, Perucho M, Kunkel TA (1994) Defective mismatch repair in extracts of colorectal and endometrial cancer cell lines exhibiting microsatellite instability. J Biol Chem 269:1–4

Umar A, Buermeyer AB, Simon J, Thomas DC, Clark AB, Liskay RM, Kunkel TA (1996) Requirement for PCNA in DNA mismatch repair at a step preceding DNA synthesis. Cell 87:505–514

Varlet I, Canard B, Brooks P, Cerovic G, Radman M (1996) Mismatch repair in *Xenopus* egg extracts: DNA strand breaks act as signals rather than excision points. Proc Natl Acad Sci USA 93:10156–10161

Vulic M, Dionisio F, Taddei F, Radman M (1997) Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc Natl Acad Sci USA 94:9763–9767

De Wind N, Dekker M, Berns A, Radman M, te Riele H (1995) Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. Cell 82:321–330

Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol Evol 11:367–370

Yeager Strassen N, Logsdon JM, Vora GJ, Offenberg HH, Palmer JD, Zolan ME (1997) Isolation and characterization of *RAD51* orthologs from *Coprinus cinereus* and *Lycopersicon esculentum*, and phylogenetic analysis of eukaryotic *recA* homologs. Curr Genet 31:144–157